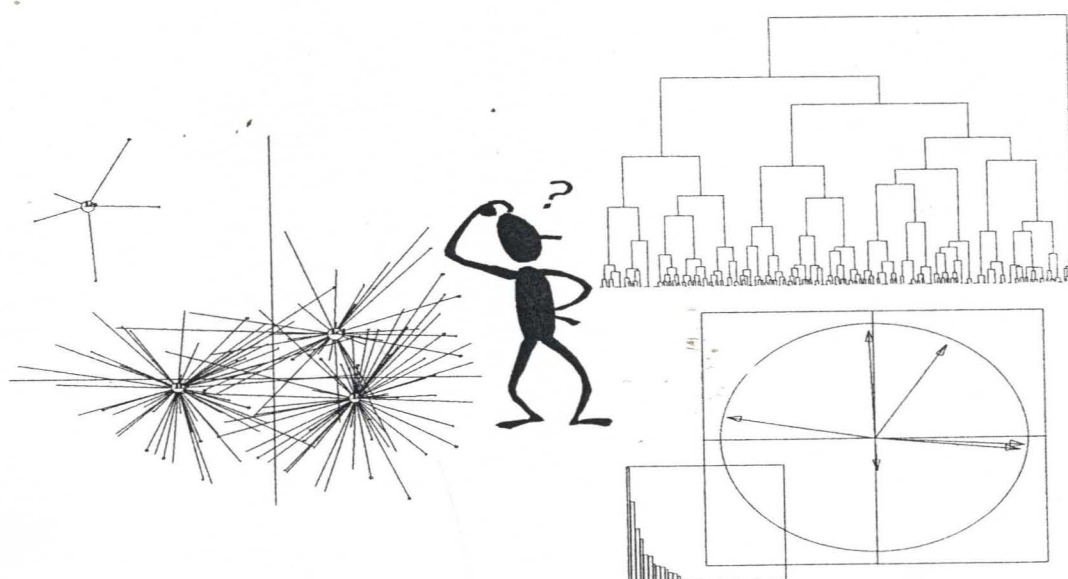




**PÔLE RÉGIONAL DE RECHERCHE APPLIQUÉE
AU DÉVELOPPEMENT DES SAVANES
D'AFRIQUE CENTRALE**

RAPPORT DE MISSION

**"SESSIONS DE FORMATION A L'ANALYSE DES DONNEES
D'ENQUETES DES CHERCHEURS PRASAC
DU CAMEROUN, DE CENTRAFRIQUE ET DU TCHAD"**



**Samir MESSAD
CIRAD-EMVT**



ICRA, IRAD, ITRAD, LRVZ,
CIRAD, IRD, Université de Leyde

AUTEUR(S) : Samir MESSAD

ACCES AU DOCUMENT :

au service de documentation du CIRAD-EMVT

ORGANISME AUTEUR : CIRAD-EMVT

ACCES A LA REFERENCE DU DOCUMENT : Libre

ETUDE FINANCEE PAR : PRASAC

REFERENCE : OM n° 30 06 01 145

AU PROFIT DE : PRASAC (Pôle régional de Recherche Appliquée au Développement des Savanes d'Afrique Centrale).

TITRE : Sessions de formation à l'analyse des données d'enquêtes des chercheurs PRASAC du Cameroun, de Centrafrique et du Tchad. Rapport de mission.

TYPE D'APPROCHE : Mission de formation.

DATE ET LIEU DE PUBLICATION : Mars 2001, Montpellier (France).

PAYS OU REGIONS CONCERNES : Cameroun, RCA, Tchad.

MOTS CLES : Formation, analyse des données, analyses factorielles, enquête, exploitations agricoles, élevage.

RESUME :

Cette mission visait à assurer la formation des chercheurs et techniciens impliqués dans la gestion et le traitement statistique des enquêtes menées dans le cadre du PRASAC (Tchad, Cameroun, RCA). Cette nécessité identifiée depuis longtemps avait en particulier été rappelée par B. Faye suite à sa mission d'appui à la composante C4 du PRASAC (rapport Cirad, n°2000-047). Deux sessions de formation à l'analyse des données ont été organisées au Cameroun et au Tchad où les participants de chacun des pays ont été regroupés par thème d'intérêt (« Exploitations agricoles » et « Elevage ») ce qui a permis de limiter la taille des groupes. Ces choix ont contribué au bon déroulement et à l'efficacité de cette formation. Nous avons fait le choix d'une présentation « non mathématique » des méthodes d'analyses des données d'enquête, méthodes qui nous le pensons, constituent une bonne porte d'entrée pour l'acquisition d'autres méthodologies statistiques. Nous espérons ainsi que les participants auront acquis après ce stage une vision plus large de la démarche statistique et de l'utilité/utilisation de ses méthodes.

Sommaire

Remerciements	3
Personnes rencontrées	3
Déroulement de la mission.....	4
Introduction	5
Liste des participants des sessions.....	6
Déroulement des sessions de formation	6
Conclusion	9
Bilan.....	9
Evaluation	9
Recommandations	10
Bibliographie	11
Annexes	13

Remerciements

Je tiens à remercier l'ensemble des personnes qui m'ont accueilli et fait en sorte que mon séjour au Cameroun et au Tchad se déroule dans les meilleures conditions. Remerciements particuliers aux organisateurs : Lamine Seyni Boukar et Jean-Yves Jamin au Tchad, Aboubakar Njoya au Cameroun, ainsi qu'à Michel Havard, à la famille Duteurtre, aux chauffeurs du PRASAC et enfin aux participants pour leur écoute attentive et intéressée.

Personnes rencontrées

Dr. Lamine Seyni Boukar, Coordinateur général du PRASAC, à N'Djamena (Tchad)

Dr. Aboubakar Njoya, Délégué national du PRASAC au Cameroun

Jean-Yves Jamin, Cirad-Tera, Coordinateur scientifique du projet PRASAC

Didier Chavatte, Conseiller technique principal à la coordination régionale du PRASAC

Déroulement de la mission

10/03/2001	Départ de Montpellier
11/03/2001	Arrivée à N'Djamena (Tchad) Rencontre avec Seyni Boukar (Coordinateur régional PRASAC) Déjeuner avec G. Duteurtre (chercheur CIRAD) et Eric Gozé (co- formateur pour le groupe « Agronomie ») Départ en voiture pour Garoua (via Maroua, Cameroun).
12/03/2001	Rencontre avec Aboubakar Njoya (Délégué National PRASAC au Cameroun), Michel Havard (chercheur CIRAD-TERA) Préparation et début de la formation groupe « Exploitations Agricoles »
16/03/2001	Séance de travail avec MBETID-BESSANE Emmanuel et David KADEKOY.
17/03/2001	Fin de la session de formation au Cameroun. Bilan avec les stagiaires.
18/03/2001	Départ en voiture pour N'Djamena.
19/03/2001	Début de la session de formation groupe « Elevage » à N'Djamena au Laboratoire de Recherches Zootechniques et Vétérinaires de Farcha (LRVZ).
20/03/2001	Séance de travail avec l'équipe « Elevage » PRASAC Cameroun Traitement statistique des données de l'enquête sur l'élevage porcin dans les provinces Nord et Extrême-Nord du Cameroun.
21/03/2001	Seconde séance de travail avec l'équipe « Elevage » PRASAC Cameroun. Traitement statistique des données de l'enquête sur les agro-pasteur du Nord et de l'Extrême-Nord du Cameroun (Enquête conduite dans le cadre du projet hydraulique pastoral).
23/03/2001	Fin de la session de formation au Tchad. Bilan avec les stagiaires et Jean-Yves Jamin.
24/03/2001	Départ de N'Djamena et arrivée à Montpellier

Introduction

La méthodologie statistique n'est plus réservée aux seuls spécialistes. Les moyens de calculs et les logiciels s'améliorant et la « démocratisation » par intégration de la Statistique dans toutes les formations scientifiques, obligent désormais tout chercheur à inclure la dimension quantitative dans ses travaux. Les études systèmes, les typologies de fonctionnement, l'identification de trajectoires d'évolution ont désormais leurs outils d'analyse, nul ne peut aujourd'hui les ignorer.

Les outils statistiques existent mais c'est véritablement une culture statistique qui fait défaut parmi les chercheurs. Il est dès lors urgent que certains d'entre eux acquièrent ou soient convaincus d'acquérir une culture statistique minimale. Les méthodes d'analyses factorielles et de classification qui ont fait l'objet de deux sessions de formation à l'intention des chercheurs PRASAC du Cameroun, de Centrafrique et du Tchad et qui sont avant tout les méthodes de bases de l'analyse des données d'enquête constituent également une bonne introduction aux autres outils statistiques et à la démarche statistique en général. C'est dans cet esprit que s'inscrit notre démarche d'initiation à l'analyse des données.

L'approche a été voulue non mathématique en privilégiant les représentations géométriques souvent plus intuitives. Nous avons cependant insisté sur les mécanismes de calcul des méthodologies, préalable indispensable à la compréhension et l'interprétation des résultats et qui nous a également permis de restituer d'autres méthodologies connues comme la régression, l'analyse de variance ou la théorie des tests statistiques. De toute évidence ces digressions ne furent pas inutiles.

Liste des participants des sessions

Groupe « Exploitations agricoles » à Garoua (Cameroun)

Nom Prénom	Pays	Module 1 ¹
KAMENI Anselme	Cameroun	oui
NTOUPKA Mama	Cameroun	oui
M'Biandoun Mathurin	Cameroun	oui
KADEKOY David	RCA	non
MBETID-BESSANE Emmanuel	RCA	non
MOUNKAMA Kenikou	Cameroun	oui
LEROY Jean	Tchad	oui
DJONNEWA André	Cameroun	oui
ABAKAR Oumar	Cameroun	oui

DJONNEWA André n'a pu suivre que les 3 premières journées (mission au Niger).
ABAKAR Oumar a été présent une seule journée.

Groupe « Elevage » à N'Djamena (Tchad)

Nom Prénom	Pays	Module 1
KOUSSOU Mian Oudanang	Tchad	non
SOUGNABE Pabame	Tchad	non
MOPATE LOGTENE Youssuf	Tchad	oui
MAHO Anyaya	Tchad	oui
LOKO DIKA Benjamin	Cameroun	oui
NDZINGU AWA Daniel	Cameroun	non
NGO TAMA Anne-Clarisse	Cameroun	oui
NJOYA Aboubakar	Cameroun	non

NJOYA Aboubakar n'a pu suivre que les 3 premières journées en raison d'obligations professionnelles au Cameroun.

Déroulement des sessions de formation

1^{ère} Journée.

Présentation du déroulement de la session et discussion avec les participants pour ce qui concerne leurs attentes et leur pratique en matière d'analyse des données.

Ci-dessous quelques thèmes cités par les stagiaires :

- problèmes de validation statistique (méthodes descriptives/méthodes inférentielles),
- pratiques d'interprétation des résultats d'analyses factorielles,

¹ Module 1 : participation au module 1 de la formation à l'analyse des données d'enquête assurée par Paulette Tailly du 20 janvier au 7 février 2001 au Cameroun et au Tchad.

- préparation des tableaux, choix des variables,
- méthodes d'analyses factorielles plus élaborées, autres logiciels disponibles.

Introduction générale sur le traitement statistique des données : l'approche exploratoire et l'approche inférentielle (modélisation) sont souvent distingués. Nous avons présenté leurs différences mais aussi leurs complémentarités. Nous avons également insisté sur l'intérêt des méthodes de l'analyse des données pour la prise en compte des phénomènes multivariés, la construction de typologie et leur pouvoir synthétique lorsque l'on dispose d'une information en quantité importante (fichier de données d'enquête notamment).

Une présentation (avantages comparatifs et complémentarités) des principaux logiciels de statistiques commerciaux ou disponibles librement sur Internet a été faite.

2^{ème} journée

Exposé des notions de base (structure et distance) à l'aide d'exemples très simples. Ces notions sont utiles à la compréhension des méthodes d'analyse factorielle et de classification mais permettent également d'appréhender l'ensemble de la méthodologie statistique.

Les méthodes factorielles sont au centre de l'analyse des données. Mise en perspective et présentation des trois méthodes de bases que sont l'ACP (Analyse en Composante Principale), l'AFC (Analyse Factorielle des correspondances) et l'AFCM (Analyse Factorielle des Correspondances Multiples) et des objectifs d'analyse qui leur sont assignés.

Revue des méthodes une à une au cours des prochaines journées en commençant par l'ACP dont le schéma d'analyse est commun aux autres méthodes.

Application pratique de l'ACP avec Winstat.

3^{ème} journée

Présentation des deux principaux types de méthodes de classification automatique : les classifications hiérarchiques et les méthodes des *nuées dynamiques*. Nous avons présenté le premier type de méthodes en disséquant les étapes de calcul de la classification ascendante hiérarchique (CAH). Introduction aux notions de décomposition de la variance et de mesure de l'effet d'un facteur qualitatif.

A l'aide d'exemples nous avons mis en évidence la complémentarité des méthodes d'analyses factorielles et de classification. En effet les objectifs dédiés sont similaires, à savoir pour les premières, synthétiser un système de relations entre variables décrivant des individus statistiques et pour les secondes, résumer à l'aide d'une partition les ressemblances entre les individus décrits par les variables. Enfin cette complémentarité peut être résumée sous forme d'un schéma de construction d'une typologie (voir annexes 1).

Application pratique avec Winstat.

4^{ème} journée

L'analyse des relations entre deux variables qualitatives à l'aide de l'analyse factorielle des correspondances (AFC). Nous avons insisté sur la distance du Khi2 comme mesure de dissemblance entre deux profils de distributions de fréquences et de dépendance entre deux caractères qualitatifs. La statistique du khi2 présentée comme statistique de test nous a conduit à une présentation sous forme d'initiation à la théorie des tests.

Application pratique avec Winstat.

5^{ème} journée

On généralise l'étude des relations entre variables qualitatives à l'aide de l'analyse factorielle des correspondances (AFCM) qui s'avère bien adaptée aux tableaux de données issues d'enquête. A l'aide d'un exemple d'enquête nous avons vu comment sont codées les données, quels sont les calculs effectués et comment interpréter les résultats numériques et graphiques. Enfin nous terminons sur le rôle joué par les individus ou variables sous statut «supplémentaire» pour l'interprétation.

Nous avons beaucoup insisté sur la transformation du tableau qualitatif en tableau disjonctif complet. Ce tableau joue un rôle clé en Statistique.

Le codage des données est une partie cruciale du traitement des données. Il conditionne dans une large mesure « l'interprétabilité » des résultats d'une analyse et le temps que l'on y consacre est important. Nous avons fait un exposé sur quelques règles de codage des variables quantitatives, insisté plus particulièrement sur les variables qualitatives qui posent plus de problèmes et avons terminé sur le traitement des données manquantes.

Evaluation de la compréhension du cours et de la formation reçue.

6^{ème} journée

Présentation des méthodes disponibles dans le logiciel de statistique multidimensionnelle Ade-4. Cette présentation a permis d'élargir la vision des stagiaires en ce qui concerne les possibilités d'analyses et de formaliser avec eux des stratégies d'approche d'un ou plusieurs tableaux de données issus d'une base de données d'enquête.

- Analyses factorielles de base à un tableau et variantes,
- Analyses factorielles à un tableau avec une relation de voisinage (spatial ou temporel),
- Analyses factorielles à un tableau avec des groupes de lignes (analyse discriminante, analyse inter/intra et tests de Monte Carlo),
- Analyses deux tableaux (co-inertie, analyses canoniques, analyses factorielles avec variables instrumentales, méthodes d'ordination partielles),
- Analyses à K-tableaux (STATIS, analyse factorielle multiple, co-inertie multiple).

Démonstration : analyse des relations d'un jeu de variables conditionnées par un facteur (analyse inter) et élimination de l'effet (analyse intra).

Conclusion

Bilan

Il faut d'abord souligner les excellentes conditions matérielles de déroulement des sessions et ce malgré les coupures d'électricité subies lors de la session à N'Djamena. La mise à disposition d'un rétro-projecteur multimédia, en permettant une interactivité plus grande avec les participants a amélioré l'efficacité des interventions du formateur.

Ensuite, l'exigence de constituer des groupes réduits (8 personnes maximum) et sensiblement homogènes du point de vue des niveaux, s'est avérée payante. La formation s'est déroulée à un rythme qui a convenu à peu près à tout le monde.

Evaluation

Afin d'avoir une idée de l'acquisition des méthodologies et de la perception du cours par les participants, nous avons procédé d'une part, à une petite évaluation qui permet d'identifier les lacunes et les éléments compris dans le cours et d'autre part, à une série de questions relatives à différents éléments d'appréciation de l'intervention (voir annexes 2).

Evaluation de la compréhension du cours

Le petit test proposé aux participants (voir annexes 2), les scinde en deux catégories : ceux pour qui il demeure encore de nombreuses confusions et ceux qui ont visiblement acquis l'essentiel qui leur permettra de revenir plus aisément sur les notions abordées (nous ne prétendons pas à ce que les stagiaires soient complètement autonomes en analyse des données à la fin de la formation).

Certaines difficultés sont liées au peu de pratique informatique en général et en particulier la manipulation de tableaux de données. Ces lacunes sont handicapantes pour une concentration optimum sur les méthodologies et la démarche car les stagiaires ont tendance alors à chercher à mémoriser machinalement l'enchaînement des manipulations du logiciel de statistique utilisé Winstat. Il faut dire également que ce logiciel a été choisi car il ne présente pas de difficultés majeures pour quelqu'un qui a l'habitude de l'interface Windows et de sa gestion des fenêtres.

Evaluation de la formation reçue

Il s'agissait pour les participants d'attribuer une note de 1 à 10 aux différents aspects suivants :

1. Correspondance de cette formation à votre attente.
2. Equilibre aspects théoriques/aspects pratiques.
3. Organisation générale du cours.

4. Qualité des supports de cours.
5. Appréciation sur la durée journalière du cours.
6. Appréciation sur la durée totale du cours.
7. Qualité pédagogique des interventions.
8. Appréciation de l'intervenant sur les aspects théoriques.
9. Appréciation de l'intervenant sur les aspects pratiques.
10. Appréciation de l'utilisation du logiciel Winstat.

Un très bon retour des participants pour la formation confirmée par les notes qu'ils ont attribuées aux différentes rubriques avec cependant les réserves classiques sur la durée totale et même la durée journalière du cours. Pour ce qui concerne la durée journalière, il s'agissait surtout de la vitesse des démonstrations d'application des analyses sous le logiciel Winstat. La plupart avaient déjà pourtant suivi la formation d'initiation à Winstat dispensée par Paulette Tailly. Nous avons plutôt effectivement longuement insisté sur les aspects méthodologiques et leur mise en application au travers d'exemples pour l'interprétation des résultats. De plus, la projection simultanée des manipulations du logiciel avec un vidéo-projecteur devait faciliter la mémorisation des enchaînements d'analyses.

Nous avons l'expérience d'une session de 2x5 jours qui ne nous a pas convaincu de l'utilité d'allonger la durée de ce type de formation. Il vaut mieux procéder en plusieurs sessions en laissant le temps au stagiaire de « digérer » et de s'éprouver, c'est à dire de confronter ses propres données aux différentes méthodologies et au logiciel.

Recommandations

Pour ce qu'il s'agit des formations en général et celle en analyse des données en particulier, il faut insister sur le fait que les méthodologies présentées au cours de ces sessions doivent faire l'objet d'une appropriation par les participants. C'est à dire qu'elle doivent être mise en œuvre par eux-mêmes dans le cadre de leur problématique de recherche. Les recettes en ce domaine n'existent pas et nous nous sommes attachés à présenter une démarche cohérente, c'est donc aux participants de faire l'effort de transposition, en espérant qu'il sera plus aisé après avoir beaucoup insisté sur les aspects opérationnels de ces méthodes en présentant des applications et exemples propres à chaque groupe de stagiaires.

Avant de parler de traitement statistique, nous avons beaucoup parlé avec les stagiaires de gestion de l'information. L'utilisation de systèmes informatisés de gestion de bases de données est encore loin d'être acquise alors qu'elle est le préalable indispensable pour un stockage, une documentation, des extractions efficaces des données faisant l'objet de traitements statistiques. Il faudrait envisager une formation autour de ces outils que sont les systèmes de gestion de bases de données relationnelles (SGBDR) pour en finir avec l'utilisation des tableurs qui sont dotés d'outils frustrés en la matière et en particulier pour la gestion des données d'enquête avec plusieurs niveaux d'information.

Le PRASAC pourrait profiter du travail initié par Michel Havard, de formalisation de la base de données de recensement des exploitations agricoles des terroirs PRASAC au Cameroun, afin de finaliser une mise en commun des informations des trois pays

partenaires (Cameroun, Tchad et RCA) sous le même support. Base de données commune qui permettrait des comparaisons aisées entre pays. Ce travail nécessiterait une compétence informatique pour ce qui concerne la formalisation (conception du MCD : Modèle conceptuel des Données) des différentes bases de données d'enquête du PRASAC.

La présence dans les équipes d'une (ou mieux, plusieurs) personne compétente en méthodologie statistique, serait un véritable capital pour les équipes du PRASAC. On pourrait envisager une formation universitaire spécialisée plutôt professionnalisante (donc moins théorique) du type DESS² qui permettrait à un chercheur ayant l'envie, les capacités et surtout le goût pour la méthodologie statistique de renforcer cette compétence véritable parent pauvre dans les centres de recherches africains. On peut également penser que cette personne jouerait le rôle de déclencheur ou d'initiateur de problématiques qui intègrent mieux la dimension statistique dans les projets de recherche.

Bibliographie

Les références citées ci-dessous ont servi à la conception du support de cours, elles sont donc essentiellement axées sur les méthodes d'analyses factorielles en privilégiant les aspects pratiques d'interprétation des résultats. Le premier titre cité constitue de notre point de vue la référence française en la matière.

[1] – Escofier B., Pagès J. (1998), Analyses factorielles simples et multiples: objectifs, méthodes et interprétation (3ème édition). Dunod, Paris.

Une référence dans le domaine, très clair même pour le débutant qui y trouvera de nombreuses indications pour l'interprétation des résultats des méthodes factorielles.

[2] – Lebart L., Morineau A., Piron M. (1995), Statistiques exploratoire multidimensionnelle, 439p. Dunod, Paris.

Très complet, il fait une présentation à la fois mathématique et pratique des méthodes factorielles simples et de classifications. Il fait aussi le point sur les méthodes récentes multi-tableaux : AFM, STATIS, etc. Son originalité tient au fait qu'il fait tenir en un seul ouvrage les méthodes statistiques « explicatives » comme la régression, l'analyse de variance, l'analyse canonique, etc. et fait le lien avec les méthodes d'analyse de données.

[3] – Bry X. (1995), Analyses factorielles simples, 111p. Economica, Paris.

Sous format poche, ce petit livre sur l'ACP et l'AFC se veut très simple et pratique. Basé sur un exemple concret, il permet de comprendre les notions et objectifs des méthodes factorielles même si l'exposé est un peu rapide. A posséder dans ses bagages pour voyager plus léger.

[4] – Fenelon J.P (1981), Qu'est-ce que l'Analyse des Données, Lefonen, Paris.

C'est le livre qui accompagne les sessions de formation de l'ADDAD en analyse des données. J.P Fenelon, l'auteur, est une figure un peu atypique en France dans ce domaine, en ce sens qu'il

² DESS Méthodes statistiques des industries agricoles, agroalimentaires et pharmaceutiques (M.S.I.A.A.P), Université Montpellier II
http://www.univ-montp2.fr/~formations/diplomes/dess/dess_methodes_stat.htm

propose une vision intuitive et non mathématique contribuant ainsi à la diffusion de ces méthodes auprès d'un public de plus en plus large.

[5] – Saporta G. (1990), Probabilités Analyse des Données et Statistiques, Technip, Paris.

Une référence avec un exposé très théorique, il réunit dans un même ouvrage la plupart des méthodes statistiques courantes, y compris inférentielles, et bien sûr les méthodes mutidimensionnelles.

[6] – Faye B. (1995), Initiation à l'analyse de données, 83p. CNPR-Ministère de l'Agriculture et de la Pêche, Lempdes.

La première partie du présent document reprend ce support de cours à l'usage des débutants en analyse des données.

[7] – Escofier B., Pagès J. (1997), Initiation aux traitements statistiques. Méthodes, méthodologie, 263p. PUR, Rennes.

C'est une présentation sous forme d'un cours des méthodes statistiques courantes pour traiter un tableau de données. Donc très didactique et pratique sans formalisme mathématique, à partir d'exemples nombreux et abordant des problèmes méthodologiques importants comme la prise en compte de données manquantes ou aberrantes.

[8] – Troude C., Lenoir R., Passouant M. (1993), Méthodes statistiques sous LISA – II – statistiques multivariées, 160p. CIRAD-SAR N°40/93.

Manuel d'utilisation du logiciel d'analyse de données LISA qui explique de manière claire comment mettre en œuvre les analyses avec le logiciel mais qui fait aussi le point de manière succincte sur les méthodes d'analyse multidimensionnelle.

[9] – Philippeau G. (1992), Comment interpréter les résultats d'une analyse en composantes principales ?, 61p. collection STAT-ITCF, ITCF.

[10] – Dervin C. (1992), Comment interpréter les résultats d'une analyse factorielle des correspondances ?, 72 p. collection STAT-ITCF, ITCF.

Ces deux documents qui se présentent sous forme de fascicule font une présentation des méthodes factorielles très orientée « utilisateur et les questions qu'ils se posent » pour ce qui l'intéresse le plus, c'est à dire l'interprétation des résultats fournis par le logiciel d'analyse de données.

[11] – Jambu M. (1989), Exploration informatique et statistique des données, 505 p., Collection technique et scientifique des télécommunications, Dunod.

Un ouvrage prenant pour point de départ les différentes formes de tableaux de données à analyser : données d'enquêtes, panels, tableaux de bord, mesures, ... et faisant le point des outils statistiques univariés et multivariés de base disponibles pour traiter ses données.

[12] – Grangé D., Lebart L. (1993), Traitements statistiques des enquêtes, 255p., Dunod, Paris.

Sans entrer dans le détail des méthodes décrites, cet ouvrage donne un bon aperçu de l'approche statistique des données d'enquêtes. Des stratégies d'échantillonnage à la mise au point du questionnaire, des analyses factorielles aux méthodes modélisatrices, les auteurs veulent montrer qu'il existe une diversité des approches et des points de vue sur des données d'enquête.

Annexes

Annexes 1. Synoptique d'une analyse typologique.

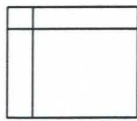
Annexes 3. Evaluation de la compréhension et de la formation reçue.

Annexes 2. Page de garde du document support de cours « Initiation à l'analyse des données ».

Annexes 3. Page de garde de la présentation « Mise au point et conception d'un questionnaire d'enquête ».

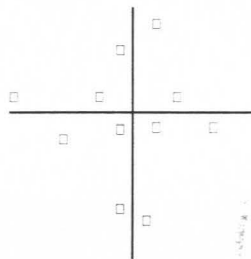
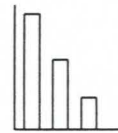
Synoptique d'une analyse typologique

Un tableau de données : des individus statistiques (lignes) décrits par des variables (colonnes).



Méthode factorielle pour décomposer la structure du tableau

indicateur de structuration du tableau (valeurs propres)



Variable 3

Variable 1
Variable 4

Variable 2

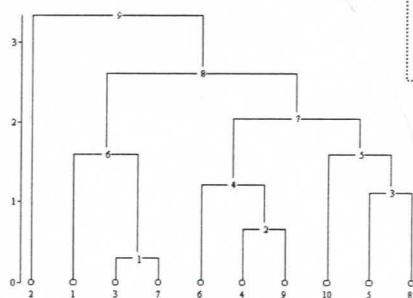
Plans factoriels

Aides pour l'interprétation :
CTR : importance de chaque variable dans la structuration du tableau
COS2 : qualité de représentation synthétique des variables

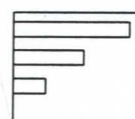
Ordonation synthétique des individus sur les facteurs

Classification à partir des coordonnées factorielles

indicateur de la partition optimale (indices de niveau)



Arbre hiérarchique



Regrouper les individus au vu de leurs similitudes (typologie)

Interprétation

Légende



Méthode



Résultat



Éléments d'appréciation

Initiation à l'analyse des données d'enquête

Merci de répondre à ce petit questionnaire qui nous permettra d'améliorer les prochaines sessions du stage « Initiation à l'analyse des données ».

Evaluation de la compréhension du cours

1. Dans une analyse factorielle, quel est l'indicateur de la structure d'un tableau de données ?

2. Soit le diagramme des indices de niveau de la hiérarchie :

3. Voici 3 tableaux. Dites quelle(s) analyse(s) doivent être utilisées ?

	type	âge	sexe
ind. 1	1	2	1
ind. 2	1	1	2
ind. 3	2	3	1

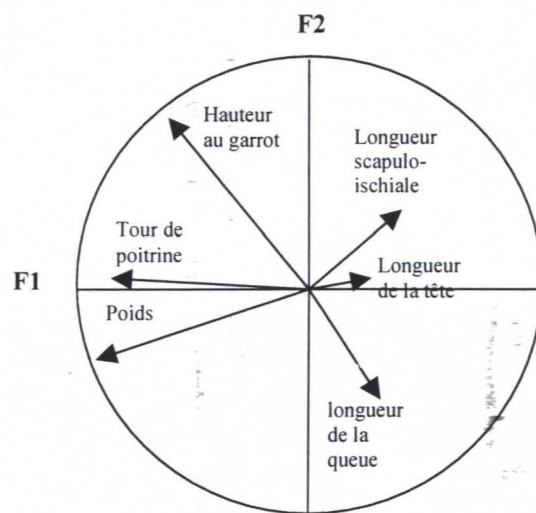
4. Soit l'analyse d'un tableau contenant 10 variables et 50 individus. Ci-joint le tableau de \cos^2 /contributions sur les 3 premiers facteurs pour 2 des 10 variables.

	F1		F2		F3	
Logement	Cos ²	CTR	Cos ²	CTR	Cos ²	CTR
<i>piquet</i>	45.6	10	2.3	0.1	34.4	12.1
<i>enclos</i>	4.6	0.5	8.6	1.9	13.9	7.3
<i>étable</i>	56.9	14.3	24.3	7.5	1.1	0.2
Alimentation						
<i>tourteau</i>	14.2	2	65	17.3	25.1	7.8
<i>fanes</i>	6.5	0.3	51.2	12.6	16.7	8.4
<i>rien</i>	2.9	0.1	1.1	0.1	0	0

- a. Quelle est la qualité de la représentation de la modalité « piquet » sur le plan F1-F2 ?
 - b. Quelle est la contribution totale des variables décrivant le logement sur le facteur F1 ?
 - c. Quelles sont les modalités les plus contributives à retenir pour chacun des facteurs ?
5. Ci-joint la représentation graphique du plan factoriel des variables d'un tableau de mesures barymétriques :
- a. Quelle analyse a-t-elle été faite ? _____
 - b. Comment s'appelle ce graphique ? _____
 - c. Quelles sont les variables expliquant les facteurs F1 et F2 ? _____

F1 : _____

F2 : _____



Evaluation de la formation reçue

Attribuez une note de 1 (très mauvais) à 10 (très bon) aux rubriques suivantes :

1. Correspondance de cette formation à votre attente.
2. Equilibre aspects théoriques/aspects pratiques.
3. Organisation générale du cours.
4. Qualité des supports de cours.
5. Appréciation sur la durée journalière du cours.
6. Appréciation sur la durée totale du cours.
7. Qualité pédagogique des interventions.
8. Appréciation de l'intervenant sur les aspects théoriques.
9. Appréciation de l'intervenant sur les aspects pratiques.
10. Appréciation de l'utilisation du logiciel Winstat.

Avez-vous des suggestions ?

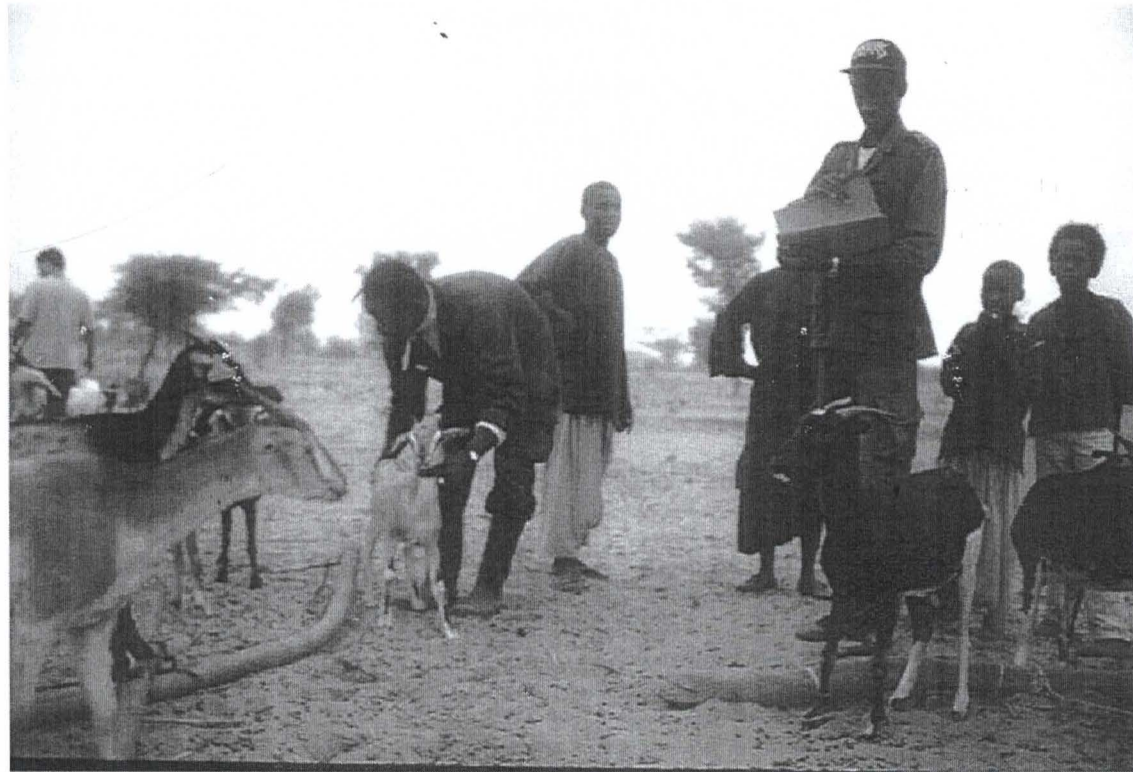
Traitement de l'information zootechnique et sanitaire

Initiation à l'analyse des données



Session de formation - CIRAD-EMVT Montpellier France

La mise au point d'un questionnaire d'enquête



B. Faye, Sénégal, 1993.

Samir Messad

CEAV - pathologie tropicale - 14 nov. 2000